

Tilburg University

Screening Experiments for Simulation

Kleijnen, J.P.C.

Publication date:
2007

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Kleijnen, J. P. C. (2007). *Screening Experiments for Simulation: A Review*. (CentER Discussion Paper; Vol. 2007-21). Operations research.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



No. 2007–21

SCREENING EXPERIMENTS FOR SIMULATION: A REVIEW

By Jack P.C. Kleijnen

March 2007

ISSN 0924-7815

Screening experiments for simulation: a review

Jack P.C. Kleijnen

*Department of Information Systems and Management, Tilburg University,
Postbox 90153, 5000 LE Tilburg, Netherlands*

Abstract

This article reviews so-called screening in simulation; i.e., it examines the search for the really important factors in experiments with simulation models that have very many factors (or inputs). The article focuses on a most efficient and effective screening method, namely Sequential Bifurcation. It ends with a discussion of possible topics for future research, and forty references for further study.

Key words: Screening, Metamodel, Response Surface, Design

JEL: C0, C1, C9

1 Introduction

Why is screening needed in simulation? The *Pareto* principle or *20-80* rule implies that only a few factors (simulation inputs) are really important (or ‘active’, as some authors say). The *parsimony* principle or *Occam’s razor* implies that a simpler explanation is preferred to a more complex explanation—all other things being equal. *Screening* means that the simulation analysts are searching for the really important factors among the many factors (often hundreds or more) that can be varied in the simulation experiment. In other words, effects are assumed to be ‘sparse’. The *curse of dimensionality* is also mentioned in many publications, including a 2002 panel report [30]. In his famous article [23], Miller claims that people cannot handle more than ‘seven plus or minus two’ factors when processing information. (Because simulation is applied in so many different disciplines, the terminology varies widely; this article includes different terms for the same concept.)

An example of a simulation with many factors is the following case study. Originally, Persson and Olhager developed a supply chain simulation for the Ericsson company in Sweden, and simulated only nine combinations of factor

values (these combinations are also called ‘scenarios’); see [25]. When we revisited this simulation model in [17], we actually distinguished 92 factors. Even if we wished to experiment with the minimum number of values per factor (namely 2), we could not simulate all combinations: $2^{92} \approx 5 \times 10^{27}$, which is close to infinity. A different design changes only one factor at a time; this design still requires 93 simulation runs if not more than two values per factor are simulated; moreover, this approach does not enable the estimation of any factor interactions; see [16]. Subsection 2.3 will show how we actually simulate only 21 combinations—each combination replicated five times—to identify a shortlist with the 11 most important factors among the original 92 factors. Note that one replicate takes 40 minutes in this case study (after modification of the simulation code, which originally took three hours per replicate).

The importance of factors depends on the *experimental domain* (also called the experimental area or experimental frame; see [39]). The users should supply information on this domain, including realistic ranges of the individual factors and limits on the admissible factor combinations (e.g., some factor values must add up to 100%). In practice, *user involvement* is therefore crucial for the application of screening methods.

There are several types of screening designs. All these designs treat the simulation as a *black box*; i.e., the simulation model transforms observable inputs into observable outputs, whereas the values of internal variables and specific functions implied by the simulation’s computer modules are unobservable.

Classic two-level designs and Frequency Domain Experimentation (FDE)—discussed in [16]—are often considered to provide screening designs. Especially so-called resolution-III designs are often called screening designs in the literature; see, for example, [14] and [38]. Another class of designs called ‘conference designs’ require twice as many combinations as there are factors; see [13]. So these designs are not practical if the simulation model is *expensive*; i.e., a single run with the model takes relatively much computer time.

Supersaturated designs have fewer combinations than factors. These designs, however, assume that the designs are not sequential, so they are relatively inefficient. Recent discussions of supersaturated designs are [1], [15], [21], [36], [37], and [40]. ([32] also gives a supersaturated design, but it uses Moving Least Squares instead of classic linear regression analysis.)

Group-screening designs aggregate (or confound) individual factors into groups so that k factors may be evaluated in less than k combinations. Consequently, these designs are also supersaturated—but they are executed in two or more steps (stages). There are several types of screening designs. Examples are One-factor-At-a-Time (OAT), Morris’s OAT, Cotter’s design, Andres’s Iterated Fractional Factorial design (IFFD), multi-stage group screening, and Se-

quential Bifurcation (SB); see [2], [4], [8], [12], [24], [26], [27], and [28]. Note that [11] gives a Bayesian analysis of screening experiments, but Bayesian approaches are not further considered in this article.

Different designs are based on different mathematical assumptions concerning the *smoothness* of the Input/Output (I/O) function implied by the underlying simulation model, possible *monotonicity* of this function, etc. This article focuses on SB because it is a very efficient and effective method if its assumptions are satisfied; see Section 2.

The *fixed* sample-size assumption of classic and supersaturated designs does not hold if the next factor or input combination is selected after the preceding simulation I/O data are analyzed. This analysis may give designs that are purely sequential, multi-stage, or two-stage (also called double sampling). Moreover, these designs are ‘customized’; i.e., they account for the specific simulation model; also see [18] and [33].

The focus of this article is on discrete-event simulation. Both random and deterministic simulations are covered in [16].

The remainder of this article is organized as follows. Section 2 presents SB. Subsection 2.1 gives an outline of the simplest type of SB. Subsection 2.2 covers some mathematical details of this simplest SB. Subsection 2.3 summarizes a case study, namely a supply-chain simulation for Ericsson in Sweden. Subsection 2.4 extends SB, accounting for two-factor interactions. Section 3 presents conclusions and possible topics for future research. Forty references are given to enable further study of screening.

2 Sequential Bifurcation

Originally, Bettonvil developed SB in his doctoral dissertation, [6]. His dissertation is summarized in [7] and [17]. A few other authors have extended SB; see [9], [10], [29], [34], and [35]; these references will be discussed below.

All screening methods assume a specific *metamodel*. A metamodel is an approximation of the I/O function that is implicitly defined by the underlying simulation model. SB uses the following metamodel assumptions, which will be spelled out below:

Assumption 1(a): a first-order polynomial is a valid metamodel.

Assumption 1(b): a first-order polynomial augmented with two-factor interactions is a valid metamodel; this assumption may replace Assumption 1(a).

Assumption 2: the signs of the first-order effects are known.

Assumption 3: there is ‘strong heredity’ if Assumption 1(b) holds.

2.1 Outline of simplest SB

The SB procedure is *sequential*. Its first step places all factors into a single group, and tests whether or not that group of factors has an important effect (this statistical test will be presented in Subsection 2.2). If the group indeed has an important effect, then the second step *splits* the group into two subgroups—*bifurcates*—and tests each of these subgroups for importance. The next steps continue in a similar way, discarding unimportant subgroups and splitting important subgroups into smaller subgroups. In the final step, all individual factors that are not in subgroups identified as unimportant, are estimated and tested.

The simplest type of SB is based on the following assumptions.

Assumption 1(a): a valid metamodel is a first-order polynomial plus noise:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + e. \quad (1)$$

The simulation input variables x_j ($j = 1, \dots, k$) are standardized such that they are either -1 or $+1$; also see [16]. This scaling implies that the factors may be ranked (sorted) by their main effects; i.e., the most important factor is the one with the largest absolute value of its first-order effect or main effect; the least important factor is the one with the effect closest to zero. The larger the range of an untransformed (original) factor is, the larger the response difference and hence the main effect of that factor is. (Also see the ‘unit cost’ effects in [10].) The noise e in (1) arises from approximation error and the use of Pseudo-Random Numbers (PRNs). If the metamodel is valid, then this noise has zero mean: $E(e) = 0$.

To estimate the parameters in the simple metamodel (1), it is most efficient to experiment with only two levels (values) per factor. In practice, it is important that these levels are realistic extreme values; so the users of the underlying simulation model should provide these values. Also see the discussion of scaling in [34].

Assumption 2: the signs of all main effects are known and are non-negative:

$$\beta_j \geq 0 \quad (j = 1, \dots, k).$$

Without Assumption 2, main effects might cancel each other. However, if

Assumption 2 holds, then the analysts can define the two levels of an individual factor such that changing the level from the standardized value -1 to $+1$ does not decrease the expected simulation output (i.e., that change either increases the output or does not change it at all). For example, if the arrival rate is increased, then the expected steady-state waiting time increases. If the queuing discipline changes from FIFO (First-In-First-Out) to SPT (Shortest-Processingtime-First), then the expected waiting time decreases; consequently, the level -1 should correspond with SPT and the level $+1$ with FIFO.

Assumption 2 is related to the *monotonicity* of the I/O function. A function $w = f(x_1, \dots, x_k)$ is called monotonically increasing if $\partial w / \partial x_j \geq 0$ for all j ($j = 1, \dots, k$). Experience shows that Assumption 2 may be easily satisfied in practice; i.e., it is straightforward to define the upper and lower level of each factor such that changing a factor from its lower to its upper level does not decrease the expected response. For example, in the Ericsson supply-chain case-study some factors refer to transportation speeds: the higher these speeds, the lower the Work In Process (WIP) and hence the lower the cost—which is the output of interest in the screening experiment. Other authors give more examples; see [20] and [22].

Note that in unconstrained optimization, the function to be maximized or minimized is usually assumed not to be monotonically increasing; otherwise, the maximum or minimum lies at the limits of the experimental area. This assumption may still be compatible with the known signs assumption; i.e., switching the (standardized) factor value from -1 to $+1$ may increase the output so this factor will be found to have an important effect. However, a ‘pathological’ counterexample is an I/O function that is not monotonic and happens to give the same output values at the two observed input levels -1 and $+1$ so the factor effect seems to be zero and this factor will be eliminated by SB.

Nevertheless, if in a particular case study Assumption 2 seems hard to meet for a few specific factors, then these factors should be treated *individually*; i.e., none of these factors should be grouped with other factors in the SB procedure. For example, [12] creates some subgroups of size one in a multi-stage group-screening design; this design is less efficient than SB, but it also uses aggregation. In general, treating such factors individually is safer than assuming negligible probability of cancellation within a subgroup.

The *efficiency* of SB—measured by the number of simulated factor combinations (and hence simulation time)—improves if the individual factors are labeled such that factors are placed in increasing order of importance; see [6], p. 44. To realize this efficiency gain, it is crucial to utilize prior knowledge of users and analysts about the real system being simulated. For example, if they conjecture that environmental factors are most important, then these factors

should be placed at the end of the list of factors. Indeed, in the Ericsson case study [17] places the environmental factor ‘demand’ at the very end of the list with 92 individual factors. The efficiency further improves when placing similar factors within the same subgroup. In the Ericsson case study, all ‘test yield’ factors are grouped together; the conjecture is that if one yield factor is unimportant, then all yield factors are likely to be unimportant too. Finally, the efficiency increases if factor subgroups are split such that the number of factors per resulting subgroup is a power of two; for example, split the first 48 factors into a subgroup of 32 ($= 2^5$) factors and a subgroup of the remaining factors. However, this splitting is not recommended if it implies splitting up a group of related factors. Anyhow, splitting a subgroup into subgroups of *equal* size (like some authors do) does not need to be optimal. Also see [6], pp. 40-43.

The way SB proceeds may be interpreted through the following *metaphor*; also see Figure 1 discussed below. Imagine a lake that is controlled by a dam. The goal of the experiment is to identify the highest (most important) rocks (actually, SB not only identifies, but also measures the height of these ‘rocks’). The dam is controlled in such a way that the level of the murky water slowly drops. Obviously, the highest rock first emerges from the water! The most-important-but-one rock turns up next; and so on. SB stops when the simulation analysts feel that all the ‘important’ factors are identified; once SB stops, the analysts know that all remaining (unidentified) factors have smaller effects than the effects of the factors that have been identified.

The aggregated effect of a given subgroup is an upper limit for the value of any individual main effect within that subgroup. If the analysts must terminate SB prematurely (for example, because their computer breaks down or their clients get impatient), then SB still allows identification of the factors with the largest main effects.

SB is extended in [35], improving the control over the type-I error rates (‘false positives’), using either a two-stage approach or a fully sequential approach. Theoretically, this control does not satisfy the classic statistical requirements concerning a prespecified experimentwise error rate and a prespecified power for the *final* results—after *all* stages have been executed. Nevertheless, the numerical results in [35] look very promising.

SB is also extended to the so-called *polytope* method in [5]. The latter method is more efficient (requiring fewer combinations to be simulated), but is also more complicated (requiring the solution of a Linear Programming or LP problem after each additional observation). Moreover this method assumes main effects only (interactions will be discussed in Subsection 2.4). Note that the LP problem arises because this method computes the Ordinary Least Squares estimate under the constraint stipulating that all regression coefficients be non-negative (see Assumption 2 above).

2.2 Mathematical details of simplest SB

To explain some mathematical details of SB, the following additional notation is used.

$w_{(j);r}$: observed simulation output with the factors 1 through j set to their high levels and the remaining factors set to their low levels, in replication r ;

$\beta_{j'-j}$: sum of main effects of factors j' through j ; so

$$\beta_{j'-j} = \sum_{h=j'}^j \beta_h. \quad (2)$$

A simple estimate (a complicated estimate is given in [5]) of this group effect based on replication r is

$$\widehat{\beta_{j'-j}; r} = \frac{w_{(j);r} - w_{(j'-1);r}}{2}. \quad (3)$$

SB starts with simulating the two most extreme scenarios: in scenario 1 all k factors are set at their low levels so $x_j = -1$; in scenario 2 all these factors are high so $x_j = 1$ ($j = 1, \dots, k$). If the metamodel in (1) is valid, then

$$E(w_{(0)}) = \beta_0 - \beta_1 - \dots - \beta_k \quad (4)$$

and

$$E(w_{(k)}) = \beta_0 + \beta_1 + \dots + \beta_k \quad (5)$$

so

$$E(w_{(k)}) - E(w_{(0)}) = 2(\beta_1 + \dots + \beta_k), \quad (6)$$

which shows that the group effect estimator defined in (3) is *unbiased*.

Likewise it follows that the individual main effect is estimated through the analogue of (3):

$$\widehat{\beta_{j;r}} = \frac{w_{(j)r} - w_{(j-1);r}}{2}. \quad (7)$$

The (say) m replicates enable the estimation of the mean and the variance for each (aggregated or individual) estimated effect. For example, (7) gives

$$\widehat{\bar{\beta}_j} = \frac{\sum_{r=1}^m \widehat{\beta_{j;r}}}{m} \text{ and } s(\widehat{\bar{\beta}_j}) = \sqrt{\frac{\sum_{r=1}^m (\widehat{\beta_{j;r}} - \widehat{\bar{\beta}_j})^2}{m(m-1)}}. \quad (8)$$

This variance estimator allows unequal response variances and Common Random Numbers (CRN); also see [16].

To *test* the importance of the estimated (either aggregated or individual) main effects statistically, SB uses a Student t statistic. Different scenarios probably produce observations with different variances, and may use CRN; see (8). SB applies a one-sided test because SB assumes that all individual main effects are nonnegative. SB uses a prespecified type-I error rate (for example, $\alpha = 0.05$) per test; i.e., SB does not adjust for multiple testing (Response Surface Methodology or RSM is also a sequential procedure that does not control the type-I and type-II error rates over the whole procedure; see [16]). However, [34] does use multiple testing procedures in SB.

To *verify* (or validate) the shortlist produced by SB, the effects of the ‘unimportant’ factors may be tested through the following two scenarios, each simulated m times:

- (i) Set all these unimportant factors at their low levels, while keeping the important factors fixed (for example, at their base levels).
- (ii) Switch all the unimportant factors to their high levels, still keeping the important factors fixed.

Obviously, these two scenarios are not used in SB if verification fixes the important factors at base values (coded as 0) that are not extreme values (coded as either -1 or 1). The difference between the outputs of these two scenarios may be tested through a Student t -statistic.

How SB proceeds sequentially is illustrated in the following case study. A formal procedure for the SB steps is given in [35].

2.3 Case study: Ericsson supply-chain simulation

The Ericsson case study distinguishes $k = 92$ factors and obtains $m = 5$ replicates. per factor combination. The first extreme scenario has all 92 factors at their low levels; its average output turns out to be $\overline{w_{(0)}} = 3,981,627$. The other extreme scenario has all factors at their high levels; its average output is $\overline{w_{(92)}} = 34,013,832$. So, the estimated group effect of all 92 factors is obtained from (2), (6), and (8), and is $\widehat{\beta_{1-92}} = (34,013,832 - 3,983,627)/2 = 15,016,102$. The standard error of this estimated group effect follows from (8), and turns out to be $s(\widehat{\beta_{1-92}}) = 42,051.18$.

Note that on hindsight, this early stage might have used fewer replicates; for example, only $m = 2$ replicates would have shown that one or more factors among the 92 factors must be important.

Figure 1 shows the successive SB steps for this case study. For example, after

the initial step with its two extreme scenarios the next step divides the current group of 92 factors into two subgroups. Into the first subgroup (in the left-hand side of the figure) we decide to place all the 79 ‘decision’ (or ‘controllable’) factors; into the other subgroup we put all 13 ‘environmental’ factors. Simulation of this scenario gives an expected output between the outputs of the preceding extreme scenarios (values are not displayed). Comparison of $\overline{w_{(79)}}$ and $\overline{w_{(0)}}$ gives $\overline{\beta_{1-79}}$. Similarly, comparison of $\overline{w_{(92)}}$ and $\overline{w_{(79)}}$ gives $\overline{\beta_{80-92}}$. So, this step splits the total effect $\overline{\beta_{1-92}}$ into its two additive components. This step decreases the upper limit for any individual effect in the first subgroup and the second subgroup respectively.

SB does not split a subgroup any further when its estimated aggregated main effect is nonsignificantly positive; for example, the estimated aggregated main effect of factors 50 through 79 turns out to be negative; see again Figure 1.

In this case study, SB stops after 21 steps. The upper limit for the main effect of any remaining individual factor is then reduced to 87,759 (this value is not displayed in the figure). The SB shortlist has 11 factors; the most important factor is factor 92. Subsection 2.1 mentioned that the SB efficiency improves when factors are labeled from least important to most important; indeed, factor 92 turns out to be the most important factor and no factor labelled smaller than 43 is declared to be important. This figure also shows that the most important individual factor (namely, factor 92) has already been identified and estimated after only ten steps; the next important factor (namely, factor 49) is identified after 16 observations.

2.4 SB with two-factor interactions

This subsection summarizes SB for situations in which Assumption 1(a) is replaced by Assumptions 1(b) and 3.

Assumption 1(b): a valid metamodel is a first-order polynomial augmented with two-factor interactions and noise:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \beta_{1;2} x_1 x_2 + \dots + \beta_{k-1;k} x_{k-1} x_k + e. \quad (9)$$

Assumption 3: if a factor has no important main effect, then this factor does not interact with any other factor.

Assumption 3 is called the *strong heredity* assumption; see [36] and also [26]. Strong heredity is related to *functional marginality*, which was recently discussed in [31].

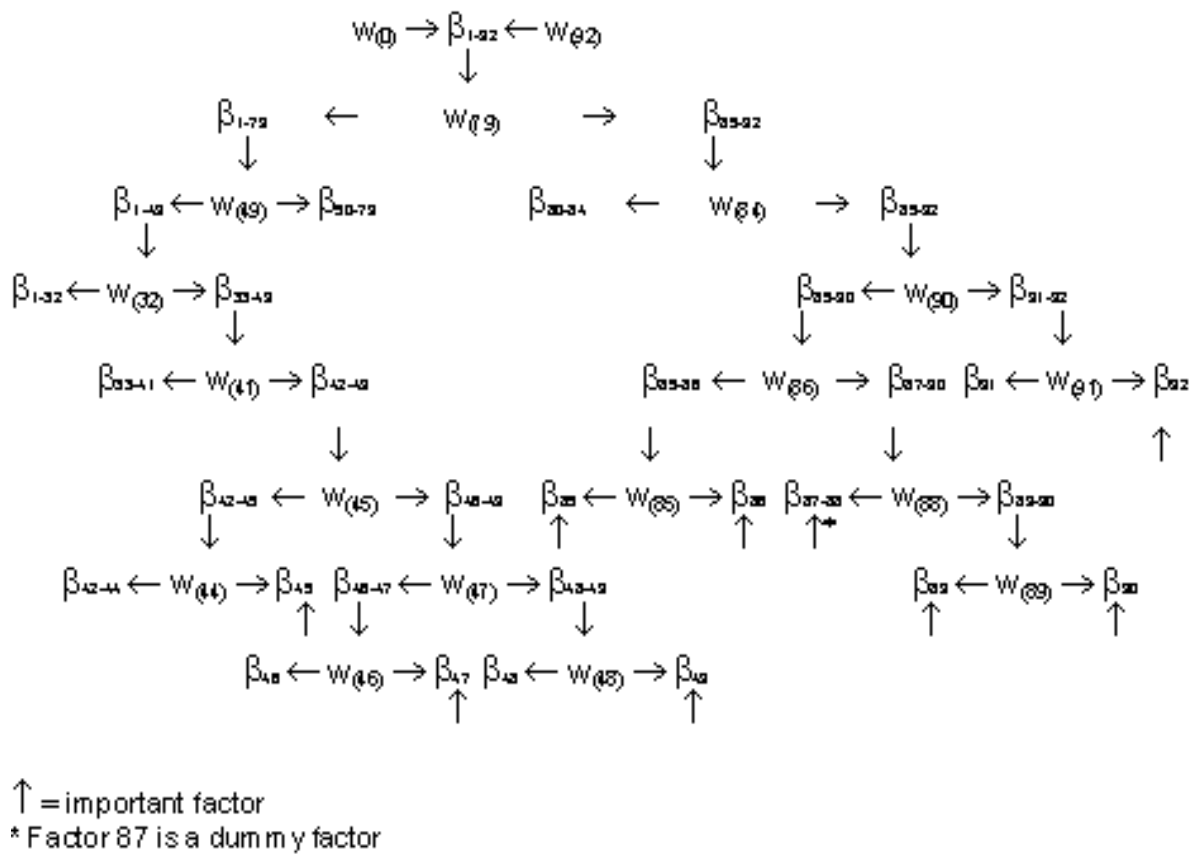


Fig. 1. SB steps in Ericsson case study

Let $w_{-(j)}$ denote the *mirror* observation of $w_{(j)}$; i.e., $w_{-(j)}$ is the simulation output with the factors 1 through j set to their low levels and the remaining factors set to their high levels. For example for $j = 48$ the analogue of (4) and (5) gives

$$\begin{aligned} E(w_{-(49)}) = & \beta_0 + (-\beta_1 - \dots - \beta_{49}) + (\beta_{50} + \dots + \beta_{92}) + \\ & + (\beta_{1;2} + \dots + \beta_{48;49}) + \\ & + (-\beta_{1;50} - \dots - \beta_{49;92}) + \\ & + (\beta_{50;51} + \dots + \beta_{91;92}) \end{aligned}$$

and

$$\begin{aligned} E(w_{(49)}) = & \beta_0 + (\beta_1 + \dots + \beta_{49}) + (-\beta_{50} - \dots - \beta_{92}) + \\ & + (\beta_{1;2} + \dots + \beta_{48;49}) + \\ & + (-\beta_{1;50} - \dots - \beta_{49;92}) + \\ & + (\beta_{50;51} + \dots + \beta_{91;92}) \end{aligned}$$

so subtracting these two equations cancels all interactions!

The analogue of (3) gives the unbiased group estimator

$$\widehat{\beta_{j'-j};r} = \frac{(w_{(j);r} - w_{-(j);r}) - (w_{(j'-1);r} - w_{-(j'-1);r})}{4}. \quad (10)$$

The analogue of (7) gives the unbiased individual estimator

$$\widehat{\beta_j;r} = \frac{(w_{(j);r} - w_{-(j);r}) - (w_{(j-1);r} - w_{-(j-1);r})}{4}. \quad (11)$$

In other words, SB enables the estimation of first-order effects unbiased by two-factor interactions if SB simulates the mirror combinations besides the original factor combinations. Hence, the number of simulated combinations doubles. However, it may happen that fewer replications per scenario are needed; see [34]. To further improve the efficiency, CRN may be applied separately to all positive levels and negative levels respectively; see [35].

SB augmented with mirror scenarios may still give misleading results if (say) two factors have unimportant main effects but their interaction is important. Therefore SB assumes *strong heredity*. If the analysts suspect that this assumption is violated for a specific factor, then they should investigate that factor after the screening phase.

SB with mirror observations does not enable estimation of *individual* interactions, but it does show whether interactions are important, as follows. Estimate the main effects from the original scenarios ignoring the mirror scenarios. If

the analyses of the mirror observations and of the original observations give the same conclusions, then interactions are unimportant. This happened, for example, in the ecological simulation reported in [6] and [7]. In that study, the factor values change relatively little (larger changes give unrealistic simulation output), so a first-order polynomial is adequate. In the Ericsson case study, however, interactions turn out to be important. (In a follow-up experiment with the factors declared to be important in SB, the sizes of the individual interactions are estimated from a Resolution V design; see [16].) The mirror observations may give a different path through the list of individual factors; for example, the path in Figure 1 may change.

3 Conclusions and future research

This article may be summarized as follows. There are different screening designs, including resolution-III, supersaturated, and group-screening designs. However, this article focused on SB, and discussed the various assumptions of SB. These assumptions may not be too restrictive in practice, as the Ericsson case study illustrated. If the assumptions are satisfied, then SB is a most efficient and effective screening method!

There is a need for more research.

It is a challenge to derive the number of replicates that control the *overall* probability of correctly classifying the individual factors as important or unimportant. SB applies a statistical test to each subgroup individually. Furthermore, SB may terminate ‘prematurely’, and yet estimate the most important factors—instead of classifying all factors with effects that exceed a prespecified threshold.

The simulation of *mirror* factor combinations may be stopped in SB as soon as it seems that no interactions are important.

After SB stops, the resulting shortlist of important factors should be *validated*. (A procedure was proposed above.)

Software needs to be developed that implements sequential screening of simulation experiments. This software should generate an input file, once a particular design type (for example, SB) has been chosen. Such a file can then be executed sequentially (and efficiently) in batch mode; that is, no human intervention is required while the computer executes the sequential design (including rules for selecting the next design point, based on all preceding observations). Good computer programming avoids fixing the inputs at specific numerical values within the code; instead, the computer reads input values

so that the program can be run for many combinations of these values. (Of course, the computer should check whether these values are admissible; that is, do these combinations fall within the experimental domain?) Such a practice can automatically provide a long list of potential factors

Multivariate output may consist of univariate outputs that require different SB paths. This problem has not yet been touched in the literature.

References

- [1] Allen, T. T. and M. Bernshteyn (2003), Supersaturated designs that maximize the probability of finding the active factors, *Technometrics*, 45, no. 1, pp. 1-8
- [2] Andres, T.H. (1997), Sampling methods and sensitivity analysis for large parameter sets. *Journal of Statistical Computation and Simulation*, 57; no. 1/4, pp. 77-110
- [3] Andres, T. (2002), Uncertainty analysis guide. Report AECL-12103, Whiteshell Laboratories, Pinawa, Manitoba R0E 1L0, Canada
<http://www.aecl.ca/Canada/aecl%20library%20site%20Canada/key%20files/serrepaeccl.htm>
- [4] Andres, T.H. and W.C. Hajas (1993), Using iterated fractional factorial design to screen parameters in sensitivity analysis of a probabilistic risk assessment model, *Proceedings of the Joint International Conference on Mathematical Methods and Supercomputing in Nuclear Applications*, Karlsruhe, Germany, April 19 to 23, volume 2, pp. 328-37
- [5] Ankenman, B.E., R.C.H. Cheng, and S.M. Lewis (2006), A polytope method for estimating factor main effects efficiently. *Proceedings of the 2006 Winter Simulation Conference*, edited by L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, pp. 369-375
- [6] Bettonvil, B. (1990), *Detection of important factors by sequential bifurcation*. Ph.D. dissertation, Tilburg University Press, Tilburg
- [7] Bettonvil, B. and J.P.C. Kleijnen (1997), Searching for important factors in simulation models with many factors: sequential bifurcation. *European Journal of Operational Research*, 96, pp. 180-194
- [8] Campolongo, F., J.P.C. Kleijnen, and T. Andres (2000), Screening methods. In: *Sensitivity Analysis*, edited by A. Saltelli, K. Chan, and E.M. Scott, Wiley, Chichester (England), pp. 65-89
- [9] Cheng, R.C.H. (1997), Searching for important factors: sequential bifurcation under uncertainty. *Proceedings of the 1997 Winter Simulation Conference*, edited by S. Andradóttir, S. K.J. Healy, D.H. Withers and B.L. Nelson, pp. 275-280

- [10] Cheng, R.C.H. and Holland, W. (1999). Stochastic sequential bifurcation: Practical issues. *Proceedings of UK Simulation Society Conference*, edited by D. Al-Dabass and R.C.H. Cheng, UKSIM, Nottingham, pp. 74–80
- [11] Chipman, H. (2006), Prior distributions for Bayesian analysis of screening experiments. In: *Screening: Methods for experimentation in industry, drug discovery, and genetics*, edited by A. Dean and S. Lewis, Springer-Verlag, New York, pp. 235–267
- [12] De Vos, C, H.W. Saatkamp, M. Nielen, and R.B.M. Huirne (2006), Sensitivity analysis to evaluate the impact of uncertain factors in a scenario tree model for classical swine fever introduction. *Risk Analysis*, 26, no. 5, pp.1311–1322
- [13] Elster, C. and A. Neumaier (1995), Screening by conference designs. *Biometrika*, 82, no. 3, pp. 589–602
- [14] Georgiou, S.D. (2006), New two-variable full orthogonal designs and related experiments with linear regression models *Statistics & Probability Letters*, in press
- [15] Gilmour, S.G. (2006), Factor screening via supersaturated designs. In: *Screening: Methods for experimentation in industry, drug discovery, and genetics*, edited by A. Dean and S. Lewis, Springer-Verlag, New York, pp. 169–190
- [16] Kleijnen, J.P.C. (2007), *DASE: Design and analysis of simulation experiments*. Springer Science + Business Media
- [17] Kleijnen, J.P.C., B. Bettonvil, and F. Persson (2006), Screening for the important factors in large discrete-event simulation: sequential bifurcation and its applications. In: *Screening: Methods for experimentation in industry, drug discovery, and genetics*, edited by A. Dean and S. Lewis, Springer-Verlag, New York, pp. 287–307
- [18] Kleijnen, J.P.C. and W.C.M. van Beers (2004), Application-driven sequential designs for simulation experiments: Kriging metamodeling. *Journal of the Operational Research Society*, 55, no. 9, pp. 876–883
- [19] Law, A.M. and W.D. Kelton (2000), *Simulation modeling and analysis; third edition*. McGraw-Hill, Boston
- [20] Lewis, S.M. and Dean, A.M. (2001), Detection of interactions in experiments on large numbers of factors (including discussion). *Journal Royal Statistical Society B*, 63, pp. 633–672
- [21] Li, L. and W. Li (2005), Tabu search and perturbation methods in the construction of supersaturated designs. *American Journal of Mathematical and Management Sciences*, 25, nos. 1 & 2, pp. 189–205
- [22] Lim, E. and P.W. Glynn (2006), Simulation-based response surface computation in the presence of monotonicity. *Proceedings of the 2006 Winter Simulation Conference*, edited by L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, pp. 264–271

- [23] Miller, G.A. (1956), The magical number seven plus or minus two: some limits on our capacity for processing information. *The Psychological Review*, 63, pp. 81-97
- [24] Morris, M.D. (2006), An overview of group factor screening. In: *Screening: Methods for experimentation in industry, drug discovery, and genetics*, edited by A. Dean and S. Lewis, Springer-Verlag, New York, pp. 191-206
- [25] Persson, J. F. and J. Olhager (2002), Performance simulation of supply chain designs. *International Journal of Production Economics*, 77, pp. 231-245
- [26] Saltelli, M. Ratto, S. Tarantola, and F. Campolongo (2005), Sensitivity analysis of chemical models. *Chemical Reviews*, 105, no. 7, pp. 2811-2827
- [27] Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto (2004), *Sensitivity analysis in practice; a guide to assessing scientific models*. Wiley, Chichester, England
- [28] Schonlau, M. and W.J. Welch (2006), Screening the input factors to a computer code via Analysis of Variance and visualization. In: *Screening: Methods for experimentation in industry, drug discovery, and genetics*, edited by A. Dean and S. Lewis, Springer-Verlag, New York, pp. 308-327
- [29] Shen, H. and H. Wan (2006), A hybrid method for simulation factor screening. *Proceedings of the 2006 Winter Simulation Conference*, edited by L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, pp. 382-389
- [30] Simpson, T. W., Booker, A. J., Ghosh, D., Giunta, A. A., Koch, P. N., and Yang, R.-J. (2004), Approximation methods in multidisciplinary analysis and optimization: a panel discussion. *Structural and Multidisciplinary Optimization*, 27, no. 5, pp. 302-313
- [31] Tsai, P-W., S.G. Gilmour and R. Mead (2006), Three-level main-effects designs exploiting prior information about model uncertainty *Journal of Statistical Planning and Inference*, in press
- [32] Tu, J. and D.R. Jones (2003), Variable screening in metamodel design by cross-validated Moving Least Squares method, *Proceedings 44th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, AIAA-2003-1669, Norfolk, Virginia, April 7-10, 2003
- [33] Van Beers, W.C.M. and J.P.C. Kleijnen (2006), Customized sequential designs for random simulation experiments: Kriging metamodeling and bootstrapping. Working Paper, Tilburg University, Tilburg, Netherlands
- [34] Wan, H., Ankenman, B. E. and B.L. Nelson (2006), Controlled sequential bifurcation: a new factor- screening method for discrete-event simulation. *Operations Research*, 54, no. 4, pp. 743-755
- [35] Wan, H., B.E. Ankenman, and B.L. Nelson (2006), Simulation factor screening with controlled sequential bifurcation in the presence of interactions, Working Paper, Purdue University, West Lafayette, Indiana

- [36] Wu, C.F.J. and M. Hamada (2000), *Experiments; planning, analysis, and parameter design optimization*. Wiley, New York
- [37] Yamada,S., M. Matsui, T. Matsui, D.K.J. Lin, and T. Takahashi (2006), A general construction method for mixed-level supersaturated design *Computational Statistics & Data Analysis*, 50, no. 1, pp. 254-265
- [38] Yu, H-F. (2006), Designing a screening experiment with a reciprocal Weibull degradation rate. *Computers & Industrial Engineering*, in press
- [39] Zeigler B.P., H. Praehofer, T.G. Kim (2000), *Theory of modeling and simulation; second edition*. Academic Press, San Diego
- [40] Zhang, Q-Z., R-C. Zhang, and M-Q. Liu (2006), A method for screening active effects in supersaturated designs. *Journal of Statistical Planning and Inference*, in press